

AD-A131 636

ON THE OPTIMALITY OF INDIVIDUAL BEHAVIOR IN FIRST COME
LAST SERVED QUEUES. (U) STANFORD UNIV CA CENTER FOR
RESEARCH ON ORGANIZATIONAL EFFICI.. R HASSIN MAY 83

1/1

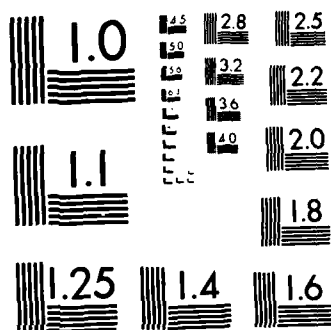
UNCLASSIFIED

TR-404 N00014-79-C-0685

F/G 12/2

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

1

ADA131036

THE OPTIMALITY OF INDIVIDUAL BEHAVIOR IN SERVING FIRST SERVED QUEUES WITH PREEMPTION AND BALKING

by
REFUEL NASSIN

TECHNICAL REPORT NO. 404
May 1983

A REPORT OF THE
CENTER FOR RESEARCH ON ORGANIZATIONAL EFFICIENCY
STANFORD UNIVERSITY

Contract Number: N00014-82-C-0001, United States Office of Naval Research

DTIC
UNCLASSIFIED
AD A131036

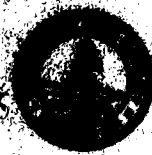
THE ECONOMICS SERIES

STANFORD UNIVERSITY STUDIES IN THE SOCIAL SCIENCES

FOURTH FLOOR, ENGINA HALL

STANFORD UNIVERSITY

STANFORD, CALIFORNIA



This document has been approved
for public release and sale; its
distribution is unlimited.

00 07 22 056

ON THE OPTIMALITY OF INDIVIDUAL BEHAVIOR IN
FIRST COME LAST SERVED QUEUES WITH PREEMPTION AND BALKING

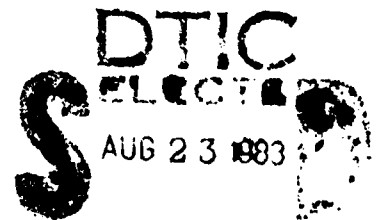
by

Refael Hassin

Technical Report No. 404

May 1983

A REPORT OF THE
CENTER FOR RESEARCH ON ORGANIZATIONAL EFFICIENCY
STANFORD UNIVERSITY
Contract ONR-N00014-79-C-0685, United States Office of Naval Research



THE ECONOMICS SERIES
INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
Fourth Floor, Encina Hall
Stanford University
Stanford, California
94305

This document has been approved
for public release and sale; its
distribution is unlimited.

ON THE OPTIMALITY OF INDIVIDUAL BEHAVIOR IN
FIRST COME LAST SERVED QUEUES WITH PREEMPTION AND BALKING*

by

Rafael Hassin**

1. Introduction

Naor, 1969, was the first to prove that individual behavior in queues is in general not socially optimal. In his paper he investigated the following queueing system: A single server facility operates on a First Come First Served (FCFS) basis with Poisson arrivals and exponential service distribution. An arriving customer may either join the end of the queue or he may choose to balk at no cost. It is assumed that a customer who balks never returns to the system. Customers are risk neutral, have identical cost per unit of service and waiting time, and receive a given identical benefit at the instant of their service completion. Since waiting time is an increasing function of the queue length, and since service time is exponentially distributed, customers' behavior is of a control limit type: There is a reservation length (equivalent to the reservation wage concept in Search Theory) such that an arriving customer joins the queue if and only if the queue is shorter than this value. Moreover, Naor's assumptions guarantee that a customer who joins the queue will not balk at a latter time since his position in the queue is secured and since the exponential distribution possesses the "memoryless property".

*This research was supported by the Office of Naval Research Grant ONR-N00014-79-C-0685 at the center for Research on Organizational Efficiency, Stanford University. The author wishes to thank Kenneth Arrow, Amihai Glazer, and Jon Strand for their helpful comments.

**Statistics Department, Tel Aviv University, Tel Aviv 69978, Israel.



A

for	<input checked="" type="checkbox"/>
on	<input type="checkbox"/>
by	<input type="checkbox"/>
codes	<input type="checkbox"/>
and/or	<input type="checkbox"/>
initial	<input type="checkbox"/>

Naor proved that the socially optimal reservation length is smaller than that resulting from individual self-optimization. Yechiali, 1971, extended this result to the more general situation where the arrival process need not be Poisson.

Two possible strategies were suggested by Naor and Yechiali in order to close the gap between the individual and the socially-optimal reservation lengths: The obvious one is to limit the queue length to its socially optimal maximum length administratively. The other is to impose tolls on customers joining the queue, of an amount equivalent to the cost of waiting m service periods, where m is the size of the gap. The collected tolls can be redistributed among all customers in the population. Assuming very large population, this redistribution will not effect the individual's decision. Edelson and Hildebrand, 1975, proved that the optimal tolls will be imposed also by a profit maximizer who can impose a two-part tariff, selling rights to check the queue with an additional toll if service is rendered (see also Levhari and Sheshinski, 1974).

In this paper we offer a third possibility, that may appear quite surprising. We claim that if the queue discipline is changed to First Come Last Served with Preemption (FCLS) then individual self-optimization is socially optimal. Under a FCLS discipline, if it is ever worth while to join the queue then it is always worth while doing so. The coming customer immediately obtains service preempting by this possible service of another customer that has arrived earlier. At this instant, the last customer in line (the one who arrived first among all customers in the system) makes the decision whether to stay or balk.

It is intuitively clear that this change in the queue discipline decreases the individual's reservation length. For, consider the customer at the end of the queue when there are L other customers in front of him. While under FCFS he has to wait $L+1$ service periods before obtaining benefit from service, now this is just a lower bound on the number of service periods that he must wait.

To see why under FCLS the individual's strategy is identical to the socially-optimal one, we first must understand why this is not the case under the FCFS discipline. In deciding whether to join the queue or balk, a customer compares his own expected waiting cost with his benefit from the service. He does not take into account the additional waiting costs he imposes on other customers by joining the queue. Therefore, he may join also when it is socially preferable for him to balk. However, this is not the case under a FCLS discipline. Also there, it is the last customer in the queue who expects the largest waiting costs and thus will be the first to balk. But in contrast to the FCFS case, regardless of whether or not this person balks, the wait of future customers is the same. This is so since all future arrivals will join the queue ahead of this customer. A decision maker who wants to maximize social benefit, compares exactly the same quantities of waiting costs and benefit, and will reach the same conclusion as would the potential balker. Note the importance of preemption, without which the customer still imposes some waiting on future customers when deciding to stay in the system.

In the following section we formally describe a model and prove the optimality of individual behavior under a FCLS discipline. For readers

to whom the preceeding arguments sound sufficiently convincing, this section can serve as an alternative derivation of the formula for the socially-optimal reservation length, as it appears in Naor's paper. The last section contains some concluding remarks concerning the possibility of administering FCLS queues and other implications of our observation.

2. Individual Optimization in a FCLS Queue

We adopt Naor's assumptions and notation:

- (i) A stationary Poisson stream of customers - with parameter λ - arrives at a single server station.
- (ii) The station renders service in such a way that the service times are independently, identically, and exponentially distributed with intensity parameter μ .
- (iii) On successful completion of service, the customer is endowed with a reward of R .
- (iv) The cost to a customer for staying in the system (either waiting or being served) is c per time unit.

We diverge from Naor's assumptions in the following ones:

- (v) A newly arrived customer joins the system and is immediately served, possibly preempting the service of another customer. At this instant each customer present in the facility (actually just the one at the end of the queue) chooses one of two alternatives: either (a) he stays in the system and incurs the losses associated with spending time in it; or (b) he balks at no additional cost, and never returns to obtain this service. The choice of one of these two alternatives will be made by the customer on comparing

the expected net gains associated with each of them. In case of a tie we assume that the customer will stay in the queue.

(vi) The queue discipline is First Come Last Served, so that the queue order is preserved while new arrivals join its head and customers leave it either by balking or when their service is completed.

Let n be the customers' reservation length, i.e., a customer balks whenever there are n other customers in front of him, including the one who is being served. Let $\rho = \lambda/\mu$, and $v(n) = [n(1 - \rho) - \rho(1 - \rho^n)](1 - \rho)^{-2}$.

We claim that if $v(n) \leq R\mu/c < v(n+1)$ then customers reservation length is n . As proved by Naor, 1969, this is also the socially-optimal reservation length. The rest of this section is devoted to a proof of this claim.

Denote by $f(i)$ the expected net benefit of the customer at the i -th position in the queue (the first in the queue is the one obtaining service). Thus, $f(i)$ is defined for $i = 1, \dots, n$.

Consider the expected net benefit of the i -th customer. The next event to occur may be either a new arrival or a termination of service. Events occur according to a Poisson process with parameter $\lambda + \mu$ - as long as the server is busy - and thus the expected waiting time for the next event is $(\lambda + \mu)^{-1}$. The next event will be a new arrival with probability $\lambda/(\lambda + \mu)$, and a termination of service with probability $\mu/(\lambda + \mu)$. This leads to the following equations:

$$f(1) = -\frac{c}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} R + \frac{\lambda}{\lambda + \mu} f(2) ,$$

$$f(i) = -\frac{c}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} f(i-1) + \frac{\lambda}{\lambda + \mu} f(i+1) \quad i = 2, \dots, n-1,$$

$$f(n) = -\frac{c}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} f(n-1).$$

The equation for $f(n)$ expresses the possibility of balking by the customer at the end of the queue.

We next prove that $f(n) \geq 0$ if and only if $R\mu/c \geq \sum_{k=1}^n k\rho^{n-k}$. Substituting $f(n) = 0$ and summing the equations for $f(i)$, $i = k, \dots, n$, we obtain

$$\sum_{i=k}^{n-1} f(i) = -\frac{n-k+1}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} \sum_{i=k-1}^{n-1} f(i) + \frac{\lambda}{\lambda + \mu} \sum_{i=k+1}^{n-1} f(i)$$

or,

$$0 = -\frac{n-k+1}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} f(k-1) - \frac{\lambda}{\lambda + \mu} f(k).$$

Therefore,

$$f(k) = -\frac{n-k+1}{\lambda} + \frac{\mu}{\lambda} f(k-1), \quad k = 2, \dots, n$$

and

$$f(1) = -\frac{n}{\lambda} + \frac{\mu}{\lambda} R.$$

The solution to these equations is given by

$$f(k) = \left(\frac{\mu}{\lambda}\right)^k R - \frac{c}{\lambda} \sum_{i=1}^k \left(\frac{\mu}{\lambda}\right)^{i-1} (n - k + 1) , \quad k = 1, \dots, n .$$

For $f(n) = 0$ we obtain

$$0 = \left(\frac{\mu}{\lambda}\right)^n R - \frac{c}{\lambda} \sum_{i=1}^n i \left(\frac{\mu}{\lambda}\right)^{i-1}$$

or,

$$\mu \frac{R}{c} = \left(\frac{\lambda}{\mu}\right)^{n-1} \sum_{i=1}^n i \left(\frac{\lambda}{\mu}\right)^{i-1} = \sum_{i=1}^n i \left(\frac{\lambda}{\mu}\right)^{n-i} .$$

Therefore, $f(n) \geq 0$ if and only if

$$\mu \frac{R}{c} \geq \sum_{i=1}^n i \rho^{n-i} = (n - \sum_{i=1}^{n-1} \rho^i)(1 - \rho)^{-1} = \frac{n}{(1 - \rho)} - \frac{\rho^n + \rho}{1 - \rho^2} = v(n) .$$

Finally, for n to be the customers' reservation length it is necessary that both $f(n) \geq 0$ and $f(n+1) < 0$. In other words, $v(n+1) > \mu(R/c) \geq v(n)$. $\frac{1}{2}$

3. Discussion

A disadvantage inherited in FCLS queues is their increased variation of waiting times. A new arrival immediately obtains service, and with probability $\mu/(\lambda + \mu)$ this service terminates without any interruption. Thus many customers do not wait at all for service, while others incur long wait. Although the expected waiting time is as in a similar FCFS queue, risk averse customers are worse off. The feeling of unfairness that this discipline

discipline arises is made stronger when balking exists. Again, the expected waiting time is as in a similar FCFS queue, but now some customers are served without waiting while others wait for a long time and finally balk. Optimality of individual behavior in a FCLS queue with balking results, among other reasons, from the model's assumption that from social point of view it does not matter who waits and who obtains service, but just how many are served and how many are waiting in the queue.

An important assumption implicit in most queueing models which allow balking, is that a balker never returns to obtain the service. A question arises about the alternatives which the customer faces. He may give up the idea of obtaining this service, or may find another facility offering similar service. In both cases, if the search cost is low, the customer may find it worth while to return to the facility at a later time in order to check its new state.^{2/} This assumption is much more crucial in FCLS queues. Here, each customer except for the one obtaining service, will benefit by balking and immediately returning to the facility. He will pretend to be a new arrival and be assigned to the head of the queue. In order to have the customers behaving in a socially-optimal manner, customers' return must be administratively prevented.

While the practical usefulness of a strict FCLS queue is doubtful, our observation has an important application concerning a much more general class of queues. When the cost per unit time of keeping certain customers queueing is high, and when these customers can be distinguished from others, it may be reasonable to give them a high priority. On the other hand, if the cost per unit queueing time is constant, it will be desirable to reduce

the overall mean queueing time by giving high priority to customers expected to have a short service time. (These ideas are discussed in a survey about the economics of queues by Levhari and Sheshinski, 1974.)

When balking is allowed, priority assignment -on a random or irrelevant basis - may be beneficial even when all customers have identical service distributions and waiting costs per unit time. Such a priority system increases customers uncertainty about the number of service periods they must wait to obtain benefit, and therefore it decreases their reservation length towards its optimal value. FCFS and FCLS are just the two extreme possibilities, while random priority systems are the intermediate cases. The best discipline depends on the customers' attitude towards risk and the cost of implementation. We note that a certain solution is particularly inexpensive; just do nothing to assure a FCFS order, and let the customers themselves take care of the discipline. In many cases the result will be far from FCFS.

A case where FCLS has negative externalities is when several queues exist for the same type of service. Suppose that customers can check upon arrival the lengths of these queues, and have to make an irrevocable decision - which queue to join. In this case, the socially-optimal rule is to join the shortest queue, and this will be the customer's choice when the service order is FCFS. However, under FCLS the coming customer is totally indifferent about what queue to join, and equilibrium is attained when each newly arrived customer chooses any of the queues with equal probabilities. 3/

Optimality of the individual behavior in FCLS queues, is preserved

also for the general arrival process analysed by Yechiali, 1971. As proved there, the optimal strategy is still characterized by a single reservation length, and we observe that since the last customer in the queue imposes no externalities, his decision is the socially optimal decision. This is not necessarily true for more general service distributions, which do not have the memoryless property. There, although the last customer imposes no externalities, other do. It is possible that from social point of view it is required that the last person in the queue will stay there and that another customer, positioned ahead of him, will balk. This may happen when the latter's residual service time is larger. However, since the individual customer does not take into account the waiting he imposes on others, the customers' behavior in this case will be different from the optimal. The same is true when the service distributions are exponential but customers have different parameter μ .

Reflection on the above discussion leads to an interesting observation. The important fact in a FCLS queue is that the person at the end of the queue remains the last one as long as he stays in the queue, and therefore imposes no externalities. This property of the queue discipline can be achieved by any policy of positioning new arrivals, as long as the newly arrived customer is never assigned to the end of the queue (unless he is the only customer there). A particularly appealing policy is to assign him to the position before the last, whenever the server is busy. This policy reduces much of the increased variation in waiting times caused by a FCLS discipline. Moreover, it reduces the customers' incentive to balk and immediately return as new arrivals; nobody, except for the last one will benefit

from doing so. Also the last customer's incentive to balk and immediately return is much weaker now. Another difficulty arises however; if the customer at the end of the queue balks, the customer positioned before him will become the last one, and all future arrivals will be positioned ahead of him. To prevent this, the later may offer the last customer a payment so that he does not balk. Such side payments must be prevented to preserve the optimal reservation length. This problem does not exist in a FCLS queue, and this can be viewed as an advantage of the discipline. Finally, we mention another advantage of FCLS queues observed by Professor Arrow: Rather than administratively forcing this discipline, we can allow the newly arriving customer to choose the position he desires in the queue!

Footnotes

- 1/ The newly arrived customer will join the system if $\mu R/c \geq v(1) = 1$, that is, $R \geq c/\mu$. This rule is independent of the arrival rate since the customer compares only his benefit and cost rates per unit time, knowing that in case of an arrival before his service terminates, he has the option of balking.
- 2/ As a matter of fact, if there is no search cost, and $\lambda < \mu$, then a customer who has no time preference concerning his service will keep balking and returning until he finds the server idle.
- 3/ The customer is concerned about the decisions of future arrivals and a nonuniform distribution will not be preserved in equilibrium.

References

- Edelson, N.M. and D.K. Hildebrand [1975], "Congestion Tolls for Poisson Queueing Processes," Econometrica, 43, pp.81-92.
- Levhari, D. and E. Sheshinski [1974], "The Economics of Queues: A Brief Survey," in M.S. et al.(eds) Essays on Economic Behavior under Uncertainty, pp. 195-212. North-Holland Publishing Co. New-York.
- Naor, P. [1969], "The Regulation of Queue Size by Levying Tolls," Econometrica, 37, pp. 15-24.
- Yechiali, U. [1971], "On Optimal Balking Rule and Toll Charges in the GI/M/1 Queueing Process," Operations Research, 19, pp. 349-370.

REPORTS IN THIS SERIES

- [illegible]

REPORTS IN THIS SERIES

- [illegible]

Reports in this Series

- 376. "Necessary and Sufficient Conditions for Single-Peakedness Along a Linearly Ordered Set of Policy Alternatives" by P.J. Coughlin and M.J. Hinich.
- 377. "The Role of Reputation in a Repeated Agency Problem Involving Information Transmission" by W. P. Rogerson.
- 378. "Unemployment Equilibrium with Stochastic Rationing of Supplies" by Ho-mou Wu.
- 379. "Optimal Price and Income Regulation Under Uncertainty in the Model with One Producer" by M. I. Taksar.
- 380. "On the NTU Value" by Robert J. Aumann.
- 381. "Best Invariant Estimation of a Direction Parameter with Application to Linear Functional Relationships and Factor Analysis" by T. W. Anderson, C. Stein and A. Zaman.
- 382. "Informational Equilibrium" by Robert Kast.
- 383. "Cooperative Oligopoly Equilibrium" by Mordecai Kurz.
- 384. "Reputation and Product Quality" by William P. Rogerson.
- 385. "Auditing: Perspectives from Multiperson Decision Theory" By Robert Wilson.
- 386. "Capacity Pricing" by Oren, Smith and Wilson.
- 387. "Consequentialism and Rationality in Dynamic Choice Under Uncertainty" by P.J. Hammond.
- 388. "The Structure of Wage Contracts in Repeated Agency Models" by W. P. Rogerson.
- 389. "1982 Abraham Wald Memorial Lectures, Estimating Linear Statistical Relationships by T.W. Anderson.
- 390. "Aggregates, Activities and Overheads" by W.M. Gorman.
- 391. "Double Auctions" by Robert Wilson.
- 392. "Efficiency and Fairness in the Design of Bilateral Contracts" by S. Honkapohja.
- 393. "Diagonality of Cost Allocation Prices" by L.J. Mirman and A. Neyman
- 394. "General Asset Markets, Private Capital Formation, and the Existence of Temporary Walrasian Equilibrium" by P.J. Hammond
- 395. "Asymptotic Normality of the Censored and Truncated Least Absolute Deviations Estimators" by J.L. Powell
- 396. "Dominance-Solvability and Cournot Stability" by Herve Moulin
- 397. "Managerial Incentives, Investment and Aggregate Implications" by B. Holmstrom and L. Weiss

Reports in this Series

- 398 "Generalizations of the Censored and Truncated Least Absolute Deviations Estimators" by J.L. Powell.
- 399. "Behavior Under Uncertainty and its Implications for Policy" by K.J. Arrow.
- 400. "Third-Order Efficiency of the Extended Maximum Likelihood Estimators in a Simultaneous Equation System" by K. Takeuchi and K. Morimune.
- 401. "Short-Run Analysis of Fiscal Policy in a Simple Perfect Foresight Model" by K. Judd.
- 402. "Estimation of Failure Rate From A Complete Record of Failures and a Partial Record of Non-Failures" by K. Suzuki.
- 403. "Applications of Semi-Regenerative Theory to Computations of Stationary Distributions of Markov Chains" by W.K. Grassmann and M.I. Taksar.
- 404. "On the Optimality of Individual Behavior in First Come Last Served Queues With Preemption and Balking" by Refael Hassin.

END

FILMED

9-83

DTIC